

Questions pratiques 2: l'inférence causale

I. Possibilités et limites des modèles statistiques

- Nous avons précédemment défini et décrit la régression comme une méthode statistique visant à *décrire* la relation ou ensemble de relations entre une variable dépendante particulière et une ou plusieurs variable(s) indépendante(s)

- Nous avons également souligné que le critère essentiel du modèle de régression est qu'il doit être une description adéquate (ou conforme) des relations observées entre la variable dépendante et la ou les variable(s) indépendante(s) dans les données à disposition

- Cette description (et sa qualité) est parfois tout ce qui est recherché
- Mais le plus souvent deux autres objectifs sont visés
 - a) La prédiction
 - b) L'inférence causale

a) La prédiction

- L'objectif est de développer une formule (une équation) afin d'être en mesure de faire des prédictions sur la variable dépendante, à partir des valeurs des variables indépendantes, que ces valeurs soient observées ou potentielles (dans le cas, par exemple, de projections)

- La validité des prédictions dépend pour l'essentiel du critère d'adéquation du modèle au phénomène empirique
- Si le modèle est élaboré sur la base d'un échantillon d'une population et que les prédictions portent sur cette population, leur validité est également dépendant des caractéristiques de l'échantillon

b) L'inférence causale

- Les variables indépendantes sont ici conçues comme des causes potentielles de la variable dépendante
- Dans une telle approche, le coefficient de régression est interprété comme « l'effet » d'une variable indépendante sur la variable dépendante

- Pour chaque variable indépendante, l'objectif est dès lors de déterminer si elle exerce effectivement un effet sur la variable dépendante et d'estimer la direction et la grandeur de cet effet
- Globalement, au travers de l'analyse de régression, on vise à examiner le système de causes, ou un sous-ensemble de ce système, qui détermine la variable dépendante

- Dans un tel objectif, l'adéquation du modèle aux données (sa pertinence descriptive) est toujours nécessaire, mais cela ne constitue pas le seul élément de sa validité
- Il faut en plus porter attention à
 - i. La structure des relations entre les variables indépendantes
 - ii. L'existence potentielle de facteurs « confondants »

- Soit une association statistique entre une variable indépendante x et la variable dépendante y
- Cette association est non causale (spurious) selon le degré auquel elle est la conséquence de la dépendance mutuelle de ces deux variables x et y vis-à-vis d'une variable tierce (z)
- cette variable tierce z est qualifiée de facteur confondant

Trois critères pour qu'une relation entre variables puisse être considérée comme causale

- i. Une association statistique entre les variables
- ii. Un ordre temporel approprié
- iii. L'élimination des explications alternatives

Ne pas réifier les modèles statistiques

- Le recours à la notion de cause dans l'analyse statistique est matière à débat
- A fortiori quand elle est appliquée à des données d'observation
- Son utilisation peut être tenue pour légitime, mais il faut en préciser la signification et la portée

- « Les hommes ont des revenus du travail plus élevés que les femmes parce que ces dernières (selon une explication) se concentrent dans des emplois mal rémunérés, travaillent moins d'heures, ont un niveau d'éducation moins élevé, ont des carrières professionnelles plus discontinues, sont victimes de discriminations, etc... »

- Toutes ces explications sont parfaitement raisonnables, et certaines d'entre elles au moins sont susceptibles d'un examen statistique
- On peut donc chercher à en appréhender la manifestation dans les données observées

- De plus, ces explications, ou d'autres encore, peuvent se combiner, chacune étant susceptible de rendre compte partiellement de l'écart systématique de revenus entre hommes et femmes
- Etudier conjointement ces explications potentielles est précisément ce que permet la régression multiple ou d'autres techniques appartenant à la même famille

- Il y a donc une justification à parler de cause ou d'analyse causale dans un tel cadre
- Mais par ailleurs, il faut avoir conscience que les modèles statistiques (de régression par exemple) sont des résumés descriptifs non pas des processus sociaux eux-mêmes, dans toute leur complexité, mais de ce qui résultent des processus sociaux, de ce qui en est le produit

- Ce que donnent à voir par exemple les modèles statistiques, ce sont les écarts systématiques de revenus entre hommes et femmes, et c'est dans ce cadre qu'est posée l'assertion que le genre est un facteur causal du revenu
- Mais un modèle statistique ne rend pas compte du processus social complexe par lequel le revenu des hommes et des femmes est déterminé

- Il y a ainsi un sens valide à dire que le niveau d'éducation a un effet sur le revenu, dans la mesure où le processus réel complexe par lequel le revenu des individus est déterminé est en partie conditionné par leur degré d'éducation

- Mais reconnaître ce fait et examiner son reflet dans un résumé statistique est quelque chose de différent que celle d'affirmer qu'un titre universitaire littéralement parlant accroît le revenu des individus

Dans l'analyse statistique, la statistique n'est pas tout

- Dans le cadre de l'analyse statistique d'un phénomène, par exemple les inégalités de revenus entre genres, on peut être amené à poser certaines assertions de nature causale visant à expliquer ces inégalités sans que ces assertions puissent faire l'objet d'un examen statistique (par exemple parce que les données à disposition ne le permettent pas)

- De telles assertions constituent des inobservables
- Elles doivent néanmoins aussi pouvoir être considérées comme valides
- Pour cela, elles doivent pouvoir s'appuyer sur des arguments convaincants ou plausibles
- Dans l'analyse statistique, la validité des assertions est statistique (« interne »), mais aussi « externe » et argumentative

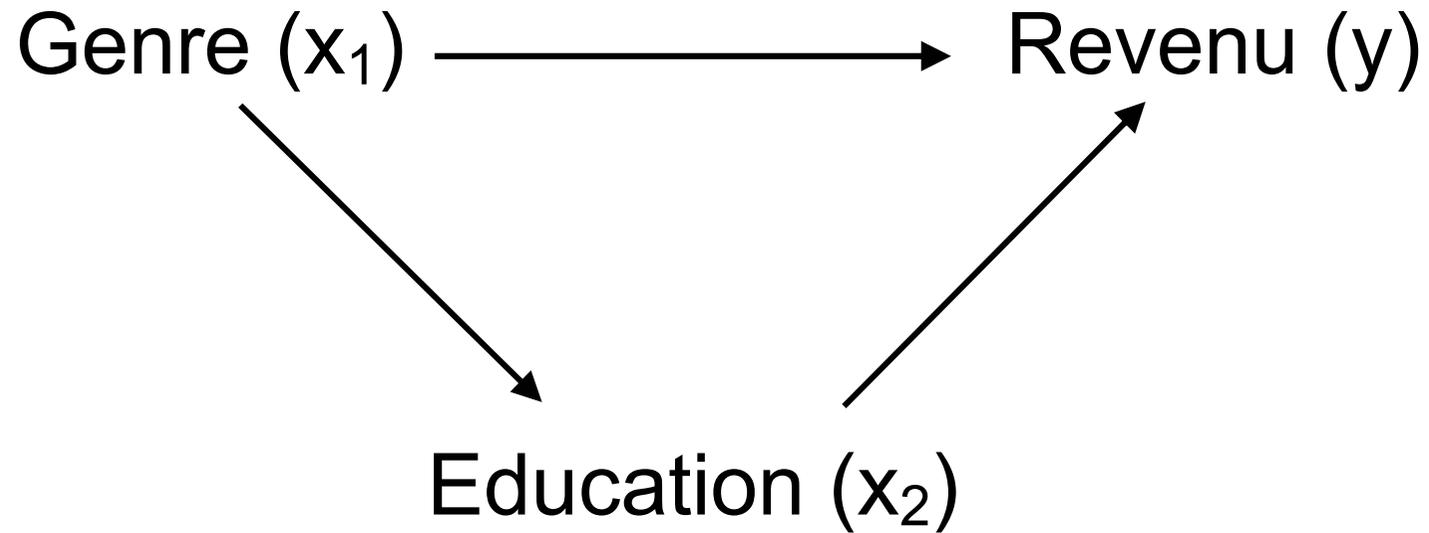
II. Variables médiatrices

- Une variable indépendante est qualifiée de médiatrice lorsqu'elle intervient comme intermédiaire dans la relation causale entre une variable indépendante et la variable dépendante

- Supposons que notre attention porte sur la relation entre genre (x_1) et revenu (y)
- L'objectif peut dès lors être celui de comprendre les mécanismes qui peuvent expliquer les écarts systématiques de revenu entre hommes et femmes

- Par exemple, il est possible que ces écarts soient entièrement ou partiellement dus au fait que les hommes ont un niveau d'éducation (x_2) plus élevé que les femmes
- Dans ce cas, il existe des différences de revenu entre genres *parce que* (entre autres) hommes et femmes ont des niveaux d'éducation inégaux
- Cf. schéma causal suivant

Exemple de schéma causal, avec éducation comme variable médiatrice



- L'éducation est ici une variable médiatrice (ou intermédiaire) de la relation causale entre genre et revenu
- Autrement dit, tout ou partie de l'effet du genre sur le revenu s'explique par le fait qu'hommes et femmes se différencient du point de vue de l'éducation *et* que l'éducation exerce un effet sur le revenu

- a) On parle *d'effet direct* pour désigner la part de la relation causale d'une variable indépendante sur une variable dépendante qui n'est pas expliquée par la ou les variables médiatrices présentes dans le modèle
- dans notre exemple, la part de l'effet causal du genre sur le revenu qui ne « transite » pas par l'éducation; représentée sur le schéma par la flèche allant directement du genre au revenu

- b) ... *d'effet indirect* pour désigner la part de la relation causale d'une variable indépendante sur une variable dépendante qui est expliquée par la présence d'une ou plusieurs variables médiatrices présentes dans le modèle
- dans notre exemple, la part de l'effet causal du genre sur le revenu qui transite par l'éducation; représentée sur le schéma par la flèche allant de genre à éducation, puis celle allant de éducation à revenu

c) ... *d'effet total* pour désigner l'addition de l'effet direct et de l'effet indirect d'une variable indépendante sur la variable dépendante

- Par définition donc:

Effet total = Effet direct + Effet indirect

- Dans notre exemple, l'effet causal total du genre sur le revenu, c'est la part de la variabilité du revenu du travail qui est due au genre ou, autrement dit, qui dépend du genre de la personne, quelque soit le ou les facteur(s) qui expliquent ce lien de dépendance; l'éducation en étant potentiellement un, pas forcément le seul

- Un modèle de régression estime toujours uniquement *l'effet direct* d'une variable indépendante, en contrôlant par toutes les autres variables indépendantes incluses dans le modèle

- Autrement dit, le coefficient de régression mesure la part de l'effet d'une variable indépendante sur la variable dépendante qui ne dépend pas des associations (corrélations) que la variable indépendante en question est susceptible d'avoir avec les autres variables indépendantes incluses dans le modèle

- Reprenons notre exemple:

Soit un premier modèle de régression (1)
ayant la forme

$$\text{revenu} = \alpha + \beta_1^1(\text{genre})$$

le coefficient β_1^1 mesure l'effet direct du
genre sur le revenu

cet effet direct est aussi l'effet total (pas
d'effet indirect)

- Incluons dans un second modèle (2) également l'éducation

$$\text{revenu} = \alpha + \beta_1^2(\text{genre}) + \beta_2^2(\text{éducation})$$

le coefficient β_1^2 mesure l'effet direct du genre, i.e. la part de l'effet qui ne dépend pas de la relation entre genre et éducation

- Autrement dit, le coefficient β_1^2 exprime la part de l'effet causal du genre sur le revenu qui ne s'explique pas par l'éducation
(qui ne s'explique pas par le fait qu'en moyenne les hommes et les femmes ont des niveaux d'éducation différents)

- De même, le coefficient β_2^2 mesure l'effet direct de l'éducation sur le revenu, i.e. la part de l'effet causal de l'éducation sur le revenu qui est indépendant de l'existence d'une association entre genre et éducation

- L'effet indirect du genre sur le revenu (i.e. l'effet du genre qui « transite » par l'éducation) peut être calculé selon la formule suivante:

$$\text{effet indirect} = \beta_1^1 - \beta_1^2$$

Autrement dit, en soustrayant de l'effet total (calculé par le modèle 1) l'effet direct qui, lui, est mesuré avec le modèle 2

- Que se passe-t-il si on fait intervenir d'autres variables susceptibles d'être médiatrices de l'effet causal du genre?
- Considérons par exemple le taux d'activité (X_3)
 - « les femmes ont un revenu du travail plus bas que les hommes parce qu'elles travaillent davantage à temps partiel »
- Et le secteur d'activité (X_4)
 - « les femmes ont un revenu du travail plus bas que les hommes parce que leurs emplois sont majoritairement dans des secteurs où les salaires sont bas »

- Le modèle de régression (3) aura la forme suivante

$$y = \alpha + \beta_1^3(x_1) + \beta_2^3(x_2) + \beta_3^3(x_3) + \beta_4^3(x_4)$$

➤ Avec

y = revenu

x_1 = genre

x_2 = éducation

x_3 = taux d'activité

x_4 = secteur d'activité

- Le coefficient β_1^3 mesure l'effet direct du genre sur le revenu, i.e. la part de l'effet causal total qui est indépendant de, ou qui ne s'explique pas par la relation entre le genre et l'éducation, ni par la relation entre genre et taux d'activité, ni celle entre genre et secteur économique

- La différence entre β_1^1 et β_1^3 représente dès lors l'effet indirect du genre sur le revenu, i.e. la part de l'effet causal total qui est expliquée globalement par l'éducation, le taux d'activité et le secteur économique

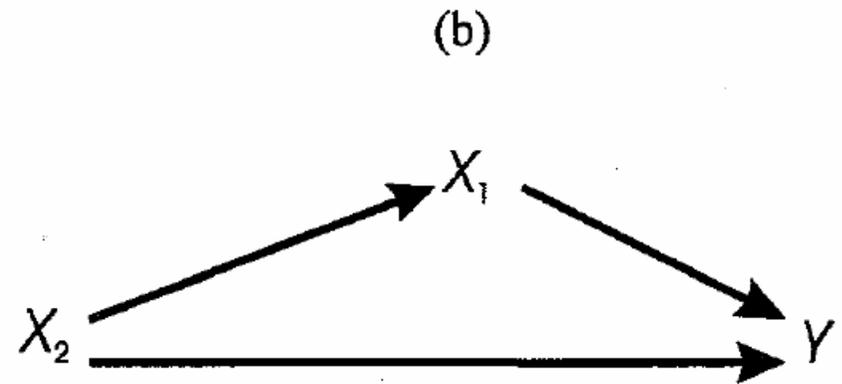
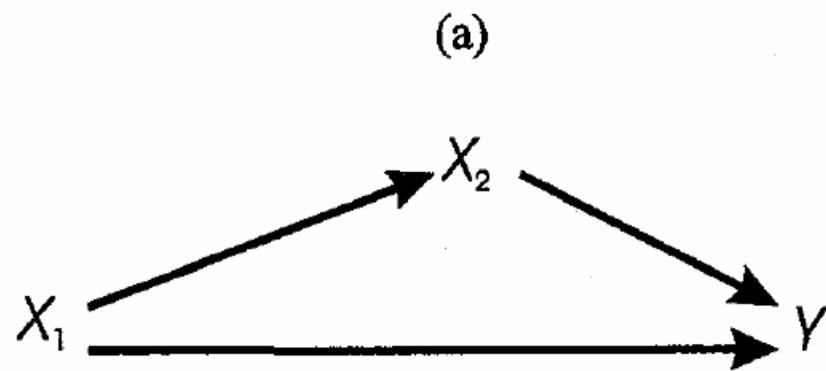
III. L'erreur de spécification

- Supposons que
 - x_1 et x_2 sont deux variables indépendantes ayant un effet causal sur y
 - x_1 et x_2 sont corrélées entre elles
- On parle d'erreur de spécification si le modèle de régression ne prend en compte qu'une des variables en question (disons x_1) et omet l'autre (x_2)

- Le valeur du coefficient de x_1 qui est calculée (B_1) est alors une estimation biaisée du « vrai » coefficient β_1
 - l'effet de x_1 inclut en effet une composante qui revient à l'effet de x_2 sur y
 - Cette composante découle directement de la corrélation (association) entre x_1 et x_2

- Pour qu'il y ait biais dû à une erreur de spécification, il faut donc à la fois
 - que le coefficient de la variable omise soit non nul (dans notre exemple $\beta_2 \neq 0$)
 - qu'il y ait une corrélation non nulle entre la variable omise et celle incluse dans le modèle (dans notre exemple $r_{12} \neq 0$)

- L'interprétation correcte de ce biais dépend de la nature de la relation entre x_1 et x_2
- Cf. les deux situations représentées dans le cliché suivant



Source: Fox (1997), p. 129.

- Dans le cas (a), x_2 est une variable médiatrice
 - L'omission de x_2 conduit à négliger le fait qu'une partie de l'effet de x_1 sur y est indirect et « transite » par x_2
 - Dans un tel cas, le coefficient β_1 réfère à l'effet total et non pas à l'effet direct de x_1 sur y
Et la valeur observée $B1$ est une estimation de cet effet total

- Dans le cas (b), x_2 est une cause antécédente à la fois de x_1 et de y
 - l'omission de x_2 conduit à attribuer à la relation empirique entre x_1 et y une composante causale qui est fausse
 - la relation entre x_1 et y étant due non à l'effet de x_1 sur y , mais à leur commune dépendance vis à vis de x_2

- Dans ce dernier cas (b), mais pas dans le premier (a), il est important de contrôler pour x_2 lorsqu'on examine la relation entre x_1 et y
- Faute de quoi, on peut être conduit à poser une conclusion fondamentalement erronée: i.e. que la relation observée entre x_1 et y est de nature causale, alors qu'elle est due au fait que x_1 et y sont tous deux dépendants de x_2 (d'une variable tierce)

Une question de degré

- Si la variable omise est faiblement corrélée aux autres variables indépendantes, ou si son effet sur la variable dépendante est de faible ampleur, le biais engendré par son omission sera peu important
- Si la variable omise est fortement corrélée à l'une ou l'autre variables indépendantes et exerce un fort effet sur la variable dépendante, le biais peut être sévère et conduire à des conclusions erronées

En synthèse

- Par rapport à une association entre une variable indépendante et une variable dépendante, l'introduction de variables indépendantes supplémentaires dites variable de contrôle, vise à:
 - i. Mettre en évidence une ou plusieurs variables médiatrices susceptibles d'expliquer cette association

- ii. Mettre à jour l'existence d'un ou plusieurs facteurs confondants, i.e. l'existence d'une composante non causale dans l'association en question

- Pour qu'une variable de contrôle soit médiatrice ou confondante, il faut à *la fois*:
 - i. qu'elle exerce un effet sur la variable dépendante y
 - ii. qu'elle soit liée (corrélée) à la variable indépendante au centre de l'attention